A neuromorphic event-based neural recording system for smart brain-machine-interfaces

Corradi, F ; Indiveri, G

DOI: https://doi.org/10.1109/TBCAS.2015.2479256

A Neuromorphic Event-based Neural Recording System for Smart Brain-Machine-Interfaces

Federico Corradi *Student Member, IEEE*, and Giacomo Indiveri *Senior Member, IEEE*

***Abstract*—Neural recording systems are a central component of Brain-Machince Interfaces (BMIs). In most of these systems the emphasis is on faithful reproduction and transmission of the recorded signal to remote systems for further processing or data analysis. Here we follow an alternative approach: we propose a neural recording system that can be directly interfaced locally to neuromorphic spiking neural processing circuits for compressing the large amounts of data recorded, carrying out signal processing and neural computation to extract relevant information, and transmitting only the low-bandwidth outcome of the processing to remote computing or actuating modules. The fabricated system includes a low-noise amplifier, a delta-modulator analog-to-digital converter, and a low-power band-pass filter. The bio-amplifier has a programmable gain of 45-54 dB, with a Root Mean Squared (RMS) input-referred noise level of 2.1 μV, and consumes 90 μW. The band-pass filter and delta-modulator circuits include asynchronous handshaking interface logic compatible with event-based communication protocols. We describe the properties of the neural recording circuits, validating them with experimental measurements, and present system-level application examples, by interfacing these circuits to a reconfigurable neuromorphic processor comprising an array of spiking neurons with plastic and dynamic synapses. The pool of neurons within the neuromorphic processor was configured to implement a recurrent neural network, and to process the events generated by the neural recording system in order to carry out pattern recognition.**

***Index Terms*—Neuromorphic system, Very Large Scale Integration, neural recordings, Very Large Scale Integration.**

I. Introduction

NEural recording systems are typically concerned with the acquisition of signals from the nervous tissue and with the transmission of these signals off-chip for further processing and analysis. As the signals being recorded are typically very small and noisy, most of the research and development efforts have been directed toward the construction of very low-noise, low-power, and high-gain amplifiers [1], [2]. To transmit the amplified signals off-chip, much research has also been dedicated to the design of wireless data links [3], [4], [5]. Thanks to these efforts, there has been tremendous progress in the development of Brain-Machince Interfaces (BMIs) that make use of these implanted microelectronic systems to reproduce as faithfully as possible the signals recorded from the neural tissue and to transmit as much of this information as possible to off-line processing stages [6], [7], [8]. The off-line computers and signal processing stages are then typically used to process the vasts amount of data being transmitted, for extracting information from interacting populations of neural cells, and for detecting action potentials, sorting them and labeling them according to the potentially multiple neurons that produced them. These systems represent extremely important tools for aiding fundamental research in neuroscience; however as the amount of electrodes for simultaneous recordings in the nervous tissue scales to very large numbers [9], [10], the energy and bandwidth required to transmit the raw data to off-chip processing stages increases to levels that are prohibitive for systems that are expected to be chronically implanted, close to the living tissue. To build neural prosthetic devices that can extract information from very large numbers of neurons and decode them *in-situ* without transmitting this information off-line, it is necessary to develop additional ultra low-power processing stages that can be interfaced to the low-noise neural signal amplifiers and integrated on the same die. In these application scenarios, it becomes important to maximize the information extracted from the raw signals, without necessarily detecting and sorting the action potentials produced by the neurons sensed by the electrodes [11].

In this paper we present a neural recording and processing system which converts the recorded bio-signals into asynchronous digital events and sends them to a low-power spiking neural network endowed with adaptive and learning abilities for decoding and classifying them on-line. We describe the features of the neural recording and signal conditioning circuits, and present demonstrations of *in-situ* signal processing using the neuromorphic architecture. The neural recording part of the system comprises a set of circuits that record, amplify, filter, and convert the bio-signals into digital asynchronous streams of pulses (see Fig. 1), which are then encoded using the AER [12]. This representation is commonly used in neuromorphic systems to implement an asynchronous communication protocol that routes and maps address-events from multiple source nodes to multiple destinations. Typically the sources of Address-Events (AEs) are pixels or nodes of neuromorphic sensory systems [13], or silicon neurons in multi-neuron chips [14]. Neuromorphic spiking neural networks can then be used to process these AEs using hardware emulations of synapses with on-line learning abilities [15], for implementing compact and low power cognitive systems that learn and adapt to the changes in the statistics of the signals being processed [16]. The goal of this work is to develop a set of circuits that can convert neural signals and act as sources of events, very much like vision sensor pixels or silicon neurons in AER neuromorphic systems, and to demonstrate learning and adaptive abilities of the neuromorphic architecture connected to them for reconstructing and classifying the recorded neural signals. The combined event-based neural recording

F. Corradi and G. Indiveri are with the Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland

Manuscript received Month DD, YEAR; revised MONTH DD, YEAR.

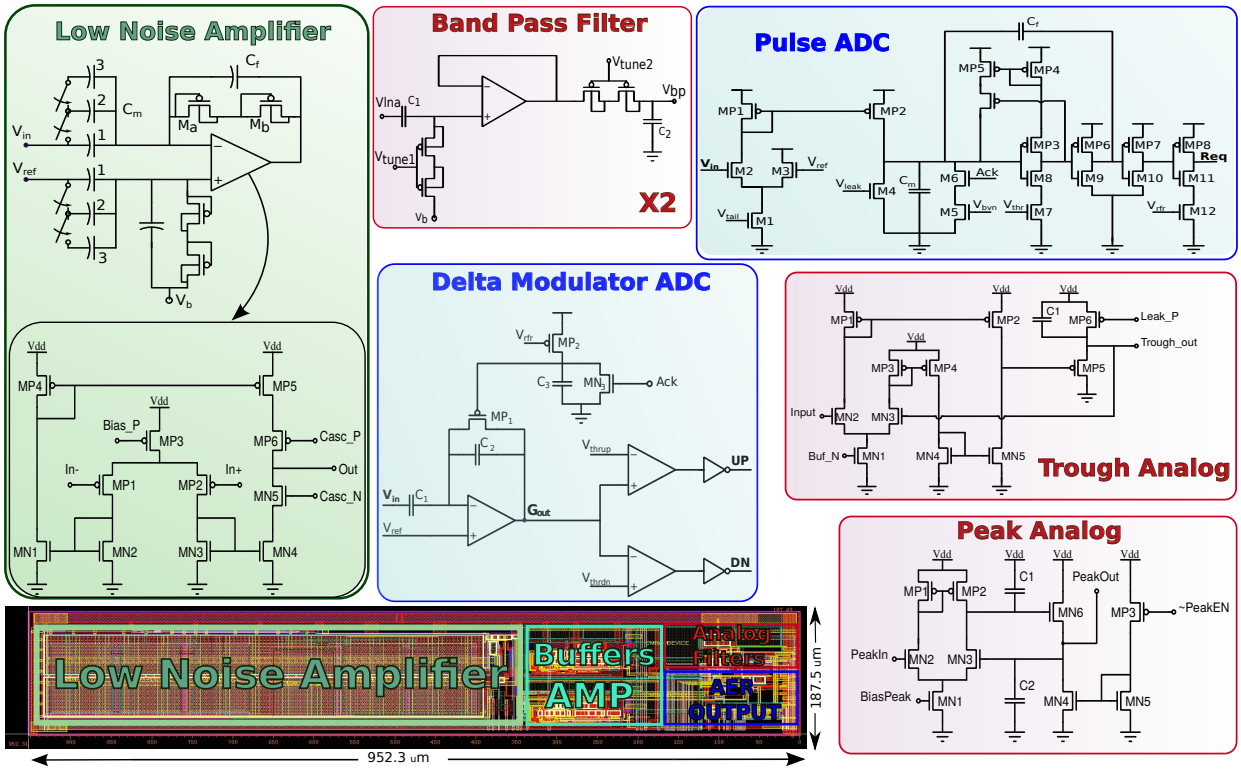


Fig. 1: Neural recording system block diagram. Bio-signals are amplified by the low-noise amplifier circuit. The output voltage of this circuit is then sent to five blocks: a) a cascade of two band-pass filters, b) an AER A/D Delta modulator, c) an analog peak detector circuit, d) an analog trough detector circuit, and e) an AER threshold-crossing spike detector circuit.

circuits and neuromorphic learning architecture represent a self-contained "smart" BMI able to produce relevant low-bandwidth control signals for prosthetic actuators, without requiring power-demanding wireless transmission of raw neural data to off-chip processing units.

The paper is organized as follow: in the next section we describe the neural recording and processing system circuits, together with the analog to digital converters. In Section III we provide experimental measurements from the neuromorphic neural architecture that characterize the response properties of all the neural recording circuits and that demonstrate the ability of the neuromorphic architecture to classify the signals produced by the neural recording circuits in an experiment in which we used real neuro-biological recordings measured form anesthetized birds exposed to two different songs, to learn to distinguish the neural responses to the two different songs. In the Discussion section we discuss about other possible neural models, alternative to the one used for improving the classification of sequences of syllables, and about alternative spiking neural network learning frameworks that can be used to decode the address-events produced by the system. In the Conclusions we summarize the achievements made and present a brief outlook for future work.

II. Methods

The overall architecture of the neural recording system integrated with the neuromorphic spiking neural network is shown in Fig. 1. The amplifier designed to record neural-, and in general bio-, signals, is a standard Low-Noise Amplifier (LNA), analogous to the one originally proposed in [17], but extended with an Analog-to-Digital Delta Modulator similar to the one proposed in [18], with event-generating circuits similar to those used in neuromorphic vision sensors [19], and with AER asynchronous communication handshaking circuits for producing the desired AEs. In parallel, the analog output of the LNA is sent to other five main blocks: a band-pass filter with pulse Analog to Digital Converter (ADC) output, an ADC delta modulator, two analog "peak" and "trough" filter circuits [20], [21], and a basic threshold-crossing spike detector circuit, to investigate potential spike-sorting capabilities of the system.

All circuits were fabricated using a standard $0.18\,\mu m$ 1-poly 6-metal Complementary Metal-Oxide-Semiconductor (CMOS) technology. The complete layout of the system occupies an area of $0.178\,mm^2$.

A. Low noise amplifier

The low noise amplifier was designed for amplifying bio-signals in the μV range, while rejecting the large *DC* component that is present at the electrode-tissue interface. Indeed, action potentials typically have amplitudes that range from $5-500\,\mu V$, depending on the distance to the electrode and on cell's size. The bandwidth of the action potential signals is in the range of $100\,Hz-3\,kHz$, whereas the Local Field Potential (LFP)s are in the frequency range $0.5-300\,Hz$.

The circuit schematic of the low noise amplifier is shown

Devices	W/L [μm]	I_d
MP3	8/4	1.3 (μA)
MP1	720/4	125 (nA)
MP2	720/4	1.25 (μA)
MN2	2/16	125 (nA)
MN1	2/16	125 (nA)
MN3	2/16	125 (μA)
MN4	2/16	125 (nA)
MN5	8/18	125 (nA)
MP6	6/16	125 (nA)
MP5	6.4/6	125 (nA)
MP4	6.4/6	125 (nA)
MP3	8/4	125 (nA)

TABLE I: Low Noise Amplifier operating points

in the first column of Fig. 1. The amplifier is a capacitive feedback circuit with mid-band gain $A_M = -C_f/C_m$, microvolt range input-referred-noise, and bandwidth $\approx g_m/(A_M C_L)$, where C_L represents the load capacitance The capacitive feedback is formed by the capacitors C_f, and C_m. The capacitor C_f connects the output of the amplifier back to its inverting input, and is in parallel with the pseudo-resistor formed by two series Metal Oxide Semiconductor Field-Effect Transistors (MOSFETs). The feedback loop controls the output signal while providing a controllable linear response. We used Metal-insulator-metal Capacitor (CMIM) capacitors of values $C_f = 35\,fF, C_{m1,m2,m3} = 15\,pF$ that are placed on top of the active area. The input switches in Fig. 1 represent transmission-gate switches that are used to set the total C_m value, resulting in a programmable gain. The differential inputs are capacitively coupled to reject the *DC* potential differences between electrode V_{in} and reference signal V_{ref}.

The main source of noise in the system is thermal noise. The MOSFET pseudo-resistors, in parallel to the capacitor C_f create the low-pass filter with a cut-off frequency at a few Hz. This design has been introduced in [2], [8], and it represents a good compromise between performances, power, and silicon area.

The very low-frequency pole is achieved thanks to the fact that the MOSFET pseudo-resistor elements (M_a, M_b in the low noise amplifier section of Fig. 1) have a much higher impedance than weak-inversion transistors, i.e., the input transistors of the amplifier (MP_1 and MP_2). To achieve low input-referred-noise it is therefore important that the input transistors work in weak inversion with microamp current levels: this can be achieved thanks to the use of very wide input transistors.

The transistor sizes for all transistors that are part of the LNA, and the DC operating point currents are shown in Table I.

B. Band-pass filter and Pulse ADC

The band-pass filter has a low cut-off frequency that depends on the properties of a Metal Oxide Semiconductor (MOS)–bipolar pseudo-resistor and the capacitor used, as described in [8]. There is however also a bias parameter that can be used to tune low-pass and high-pass cut-off frequencies. The value of this bias regulates the band of interest, for action-potentials or LFPs. The pulse ADC circuit converts the filtered signal into asynchronous spikes (see top right box of Fig. 1). The filtered signal V_{bp} is converted into a current via a differential pair operated in the weak-inversion, or sub-threshold domain [22] and controlled by its V_{ref} and the V_{tail} bias voltages. This sub-threshold current is then copied into the capacitor C_m. Constant positive offset currents can be removed via the "leak" transistor M_4 controlled by its bias voltage V_{leak}. As the voltage across the capacitor C_m increases, it slowly reaches the switching threshold of the first pseudo-inverter $M_7 - MP_4$. As the circuit starts to switch, the current flowing though it gets copied back into the C_m capacitor, therefore implementing a positive feedback loop is activated. This greatly reduces the switching time and minimizes the dissipated current. The fast switching of this pseudo-inverter triggers the generation of the Req signal, used for the acAER handshaking circuits. The pulse ADC circuit is reset by the Ack signal, coming from the AER asynchronous handshaking circuits (not shown).

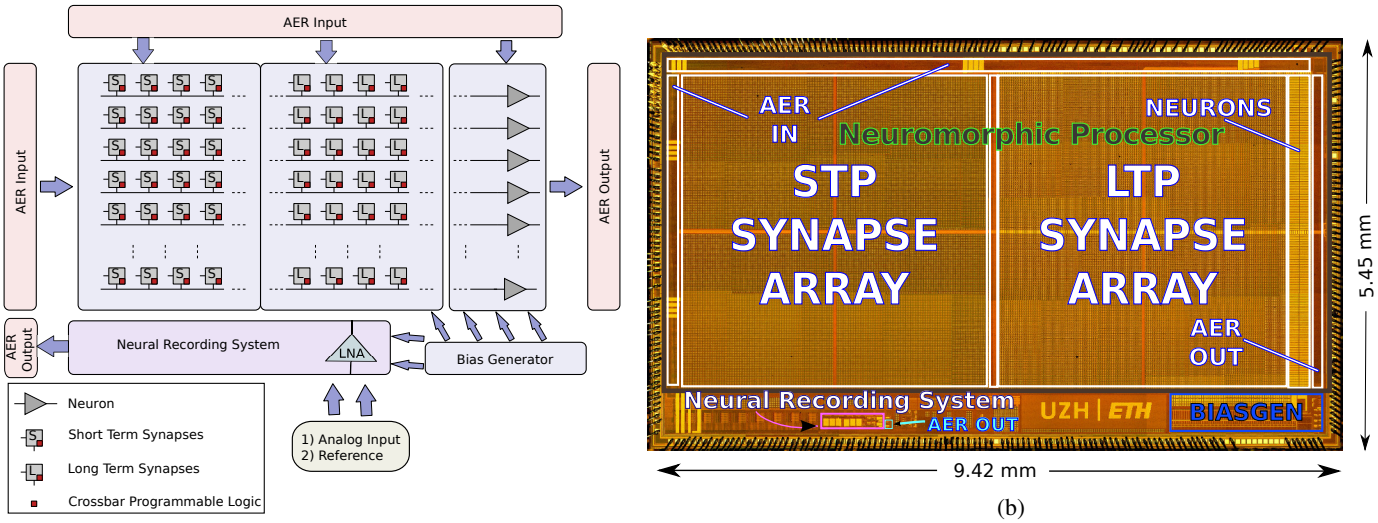
C. The asynchronous Delta Modulation A/D converter

The Delta Modulation A/D converter circuit consists of an input operational transconductance amplifier with a capacitive-divider gain stage, two comparators, and additional analog/digital circuits to manage the AER handshaking interface. Functionally, this circuit is equivalent to the one used in the Dynamic Vision Sensor (DVS) [19]: it is a self-timed clockless circuit which produces two types of digital pulses (UP or DN) if the input signal V_{in} changes by a fixed positive or negative amount respectively. The output pulses UP and DN of the circuit correspond to handshaking request signals for the AER interface. As either of these pulses are produced, the AER receiving circuits will respond with an acknowledgment signal Ack which resets the comparator output to the reference voltage V_{ref}, by shorting the amplifier output to its negative input terminal via the MP_1 p-FET. This reset state will be held for a period that is determined by the values of the C_3 capacitor and MP_2 leak current. This is essentially a "refractory period" which can be used to limit the maximum rate of AEs produced by the circuit (e.g., to control bandwidth usage).

D. The AER communication scheme

A handshaking mechanism ensures that all the digital UP and DN events generated at the sender side arrive at the receiver. These signals are encoded using a Bundled Data (BD) representation, in which the address of the channel is conveyed as a parallel word, together with two additional Ack and Req signals that are required for the handshaking control sequence. In this system time represents itself, and AEs are produced when necessary without the use of a clock signal. In the case of multiple asynchronous channels producing AEs in parallel, an arbitration block would ensure that these signals do not collide, but are transmitted on the shared bus in sequence. The throughput of multi-node/multi-channel AER systems is typically in the order of 5 megaevents/s [24].

To manage the ADCs output signals and to route the AEs produced, we used an off-chip commercial Field Programmable Gate Array (FPGA) device ($spartan-VI$). This



(a) (b)

Fig. 2: (a) Schematic diagram of the die comprising both neural amplifier and neuromorphic architecture. The neuromorphic processor contains 256 neurons, connected to 64k STP synapse circuits and 64k LTP ones. The on-chip bias generator is used to configure the analog parameters of the two systems. (b) Die micrograph. The neural recording system occupies $\approx 0.178\,mm^2$ of the entire $\sim 51.4 mm^2$ die area. (image adapted from [23]).

device represents an optimal development platform which was used mainly for data logging and AE routing.

E. The analog filters: peak detector, trough detector and level crossing

The analog filters include a peak detector circuit, a trough detector circuit and a comparator that acts as a spike detector circuit. This spike detector circuit is used to trigger the measurement of the spike amplitude. The peak and trough circuits are asymmetric voltage-followers in which the amplifier gain and the output offset voltage effect the output of these circuits. They are used to measure the peak amplitude of a spike and the deepness of the hyper-polarization of action potential signals. These features have been demonstrated to be extremely useful in spike-sorting applications [21], [25]. In the peak detector circuit, the MOSFET $MN6$ acts as a switching element, while the capacitor $C1$ is the charging element. The circuit is triggered upon the arrival of a digital *PeakEN* signal produced by a comparator circuit (not shown; it is a transconductance amplifier with a controllable bias in one of the two input terminals).

F. The reconfigurable neuromorphic processor

We used a reconfigurable spiking neural network processor [23] for carrying out neural processing tasks.

The Reconfigurable On-line Learning Spiking Neuromorphic Processor (ROLLS neuromorphic processor) of Fig. 2a contains 256 adaptive exponential integrate-and-fire neurons implemented with mixed signal analog/digital circuits. The neurons are connected to $128k$ synapses, of which $64k$ implement LTP dynamics and spike-based Hebbian-like plasticity mechanisms [26], [27], and $64k$ implement STP biologically plausible dynamics. The STP synapses do not implement learning, but are programmable with two possible weight values. In addition the STP synapses can be configured to be either excitatory or inhibitory. Thanks to additional digital logic embedded in each of the synapse blocks, the synaptic matrix allows re-configurable on-chip network connectivity, including fully recurrent networks with all-to-all connectivity. The learning mechanism of the LTP synapses is implemented thanks to spike-based pre-synaptic weight-update circuits and bi-stable long-term drift circuits [28]: while the spike-based internal dynamics of the synapse is analog, its long-term state (determined by the drift circuits) is binary. This removes the requirement of storing precise analog variables on long-time scales in Very Large Scale Integration (VLSI) and greatly simplifies the circuit design. Additional circuits are also instantiated next to the neurons array, to produce the signals that drive the weight update circuits. In Section II-G we describe the mathematical model of the learning rule that determines these signals. We refer the reader to [23] for a thorough description and characterization of the VLSI circuits.

Both the neural network connectivity state and the analog parameters of the synapse and neuron circuits are fully programmable via a high-level Python framework [29]. The combination of low-level reconfigurable hardware with the high-level Python-based programming framework allows the definition of a wide range of spiking neural network architectures, and their emulation in real-time. We used these components and features to readily explore the properties of the different neural network architectures proposed in this work and to process recorded neural signals using the real-time hardware described here for spike-pattern recognition and signal decoding tasks.

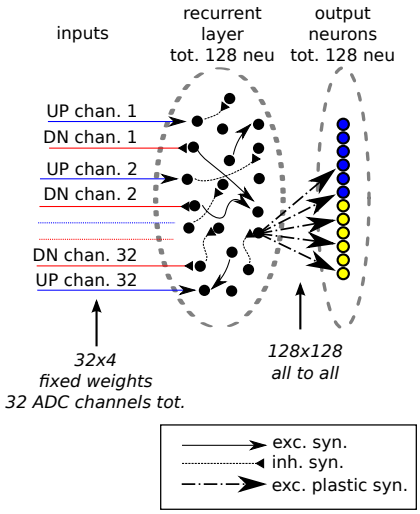


Fig. 3: On-chip network architecture. UP/DN channels represent the spiking output from the ADC delta modulator.

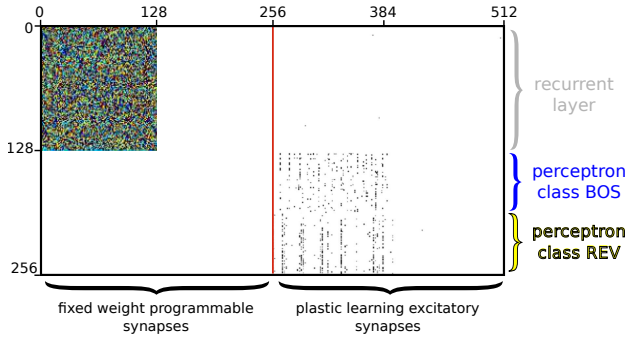


Fig. 4: Neuromorphic chip configuration. Top left quadrant: programmable STP synapse matrix. The different colors represent different synapses configurations: they can be excitatory, with either high or low weights (red or blue), or inhibitory, with weights low or high (green or yellow). Bottom right quadrant: spike-based learning LTP synapse matrix. White pixels represent synapses in their low state, black pixels represents synapses in their high state.

G. The on-chip learning rule

The LTP synapse array in the neuromorphic architecture implements a spike-based Hebbian-like learning rule [26]. The weight update rule depends on the timing of pre-synaptic spike, on the the state of the post-synaptic neuron's membrane potential, and on the recent firing history of the post-synaptic neuron. While the circuits that implement this algorithm are deterministic, there is a source of stocasticity in the pre and post-synaptic spike trains, that is used to avoid updating all the synapses in the same way. Upon arrival of a pre-synaptic spike the update rule is described by:

$$\begin{cases} w_i = w_i + \Delta w^+ & \text{if } V_{mem}(t_{pre}) > \theta_{mem} \\ & \text{and } \theta_1 < Ca(t_{pre}) < \theta_3 \\ w_i = w_i - \Delta w^- & \text{if } V_{mem}(t_{pre}) < \theta_{mem} \\ & \text{and } \theta_1 < Ca(t_{pre}) < \theta_2 \end{cases} \quad (1)$$

in which w_i is an internal analog variable that represents the synaptic weight. Δw^+, and Δw^- refer to the instantaneous up and down jumps of the internal analog variable. The post-synaptic membrane potential at the time of a pre-synaptic spike is $V_{mem}(t_{pre})$, it is compared to the threshold θ_{mem} which determines the side of the jump up, or down in the w_i changes. The post-synaptic calcium concentration is described with the term $Ca(t_{pre})$, it is proportional to the recent firing activity of the post-synaptic neuron. The parameters $\theta_1, \theta_2, \theta_3$ are thresholds that govern whether the weight is allowed to increase, decrease or should not change. Another mechanism that acts on the internal analog variable w_i is a drift mechanism. This drift is described by:

$$\begin{cases} \frac{d}{dt} w_i = +C_{drift} & \text{if } w_i > \theta_w \\ & \text{and } w_i < w_{max} \\ \frac{d}{dt} w_i = -C_{drift} & \text{if } w_i < \theta_w \\ & \text{and } w_i > w_{min} \end{cases} \quad (2)$$

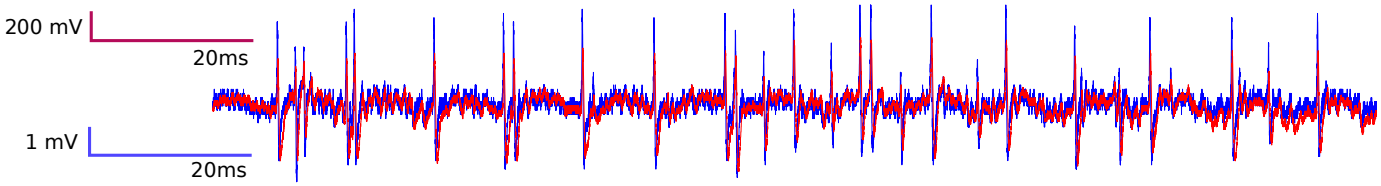
The role of this mechanism is to drive the w_i towards one of the two stable states, depending on the value of w_i compared to the threshold θ_w. The signals w_{max}, w_{min} refer to the bounds on the value of w_i. Finally, the binary weight of the synapse is a thresholded version of the variable w_i.

$$J_i = J_{max} f(w_i, \theta_J) \quad (3)$$

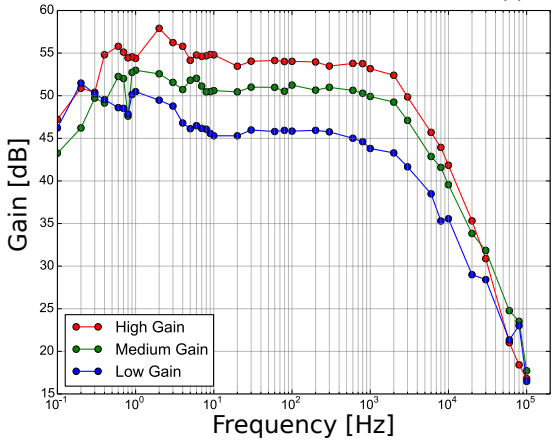
where $f(x, \theta_J)$ is a threshold function with threshold θ_J, and J_{max} is the maximum synaptic efficacy.

H. The neuromorphic classifier's architecture

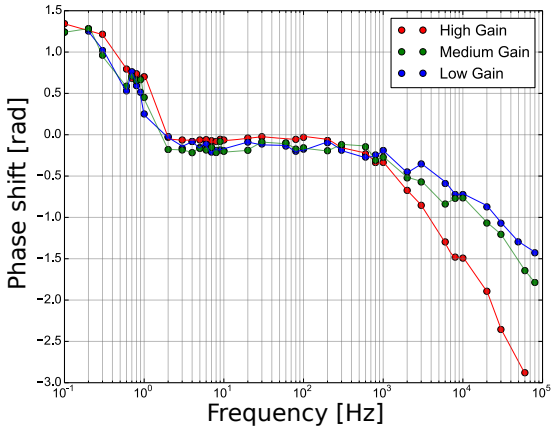
The neural core processor has been configured to implement a network of two layers, as shown in Fig. 3. The first layer of the network is composed of a recurrent pool of 128 neurons that acts as a spiking *reservoir* [30]: it receives spike-event inputs from the neural recording system, and it exploits the analog dynamics of the neurons and synapses to enhance the temporal properties of input patterns. All the neurons in the recurrent pool are connected with all-to-all projections to a second read-out layer of 128 neurons. The on-chip reservoir layer remains unchanged during training, while the connections between the first and second layer are plastic: they change their weights depending on the activity of pre and post-synaptic neurons, as described in Section II-G. In the initial configuration these synaptic contacts are reset to their low state. Figure 4 shows the configuration of the synaptic structure of the neuromorphic processor that implements the network shown in Fig. 3. The recurrent pool of neurons is implemented in the first 128 neurons that are recurrently connected via the synapses of the STP array. Every dot in Fig. 4 represents a synaptic contact. Different colors in the top left quadrant of Fig. 4 represent the different type of connections. These connections can have two possible weights and can be of two possible types: excitatory or inhibitory. The connections are initialized at random and every neuron in the recurrent pool receives on average 120 inputs, of which 85



(a) LNA Input and Output



(b) LNA Gain



(c) LNA Phase Margin

Fig. 5: a) Input and output of the Low-Noise Amplifier stage, the signal is a neural recording from a cricket (*Mecopoda Elongata*). Amplifiers is set to the minimum programmable gain (45 dB). b) Measured gain for three different values of input capacitors. c) Measured phase shift.

excitatory and 35 inhibitory. The bottom right side of Fig. 4 shows the LTP synapse matrix. Synapses in the high state are represented by black pixels. This figure highlights the state of the synapses after training; the on-chip configuration of this network topology only exploits 1/4 of the total synapses available in the device.

III. Results

A. Low noise amplifier measurements

To characterize the response properties of the LNA to signals with magnitude comparable to that of physiological ones, we applied an input sine wave produced by a Hewlett Packard 33120*A* Function Generator, setting the amplitude such that after a 100:1 resistive voltage divider the signal was 0.61 mV (the minimum amplitude range measurable by our oscilloscope). The voltage divider was a standard 2-resistor circuit, consisting of a $24.9\,k\Omega$ resistor and a $249\,\Omega$ resistor. Measurements were taken for three different gain configurations, set externally with jumpers to activate or deactivate the input capacitors shown in Fig. 1. The transfer function for the three gain settings is show in Fig. 5b. The highest gain setting gives a maximum of 58 dB, and a flat-band gain of 54 dB between 10 Hz and 2.5 kHz. The circuit still has strong amplification in the low frequency range, but the input has noticeable deformations for frequencies below 3 Hz.

As shown in Fig. 5c, the circuit provides very little phase shift in the 2 Hz-200 Hz range, and it is smaller than $-\pi/2$ rads until above 7 kHz. The gain seems to affect the phase of the output signal at high-frequency ends (see fig. 5c).

We also tested the LNA with pre-recorded signals from the auditory system of a cricket (Mecopoda Elongata), in which a single electrode was used (data courtesy of Prof. Manfred Hartbauer). The recordings were played with a standard computer audio card and attenuated to scale the signals to physiological levels. In Fig. 5a we show the result of this experiment. Since the input signal is in the millivolt range, the amplifier has been set to the minimum gain. This produces an output that is in the range of $200mv$, with no visible deformations on the amplified signal.

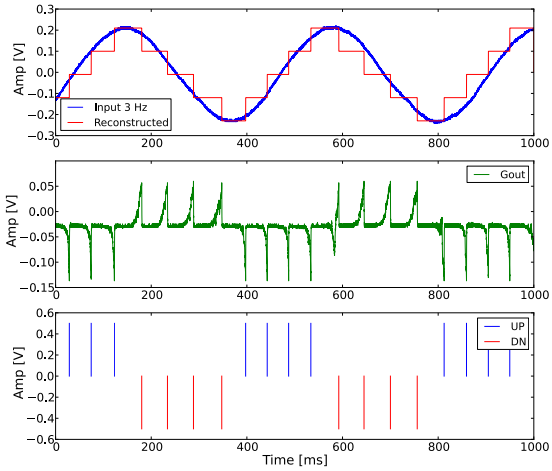
The effective Noise Efficacy Factor (NEF) achieved (see Table II) is not impressive if compared to state-of-the art amplifiers for neural recordings as in [31], [32], [33]. This is partially due to the limited bandwidth of our amplifier in comparison to state-of-the art implementations that currently achieve effective NEF 4,5 times smaller with similar input-referred-noise values.

B. The A/D asynchronous delta modulator

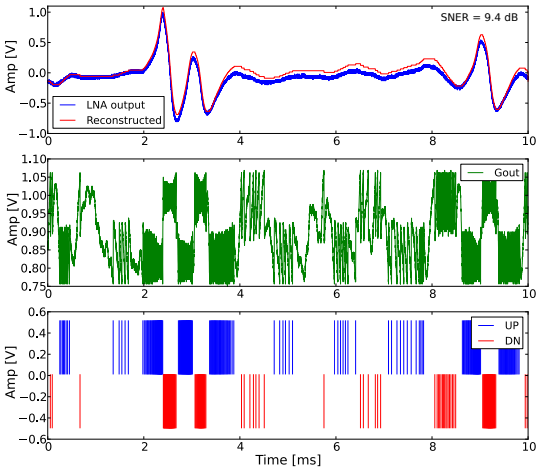
The output measured from the delta modulator circuit in response to a sine-wave input is shown in Fig. 6a. The top trace shows the output of the transconductance amplifier G_{out}, together with a reconstruction of the signal. When G_{out} exceeds one of the two thresholds V_{thrup} or V_{thrdn}, the UP pulse (third trace from the top) or DN pulse (bottom trace) are produced, and G_{out} is reset to $V_{ref} = V_{dd}/2 = 900mV$. The threshold voltages of the delta modulator are set to $V_{thrup} = V_{dd}/2 + V_{dd}/10$, $V_{thrdn} = V_{dd}/2 - V_{dd}/10$.

1) Decoding the UP and DN events: The reconstruction of the input signal from the UP and DN AEs is carried by the execution of the following algorithm:

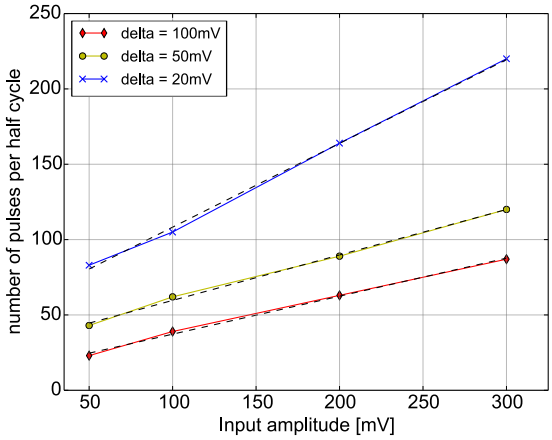
The parameters δ_{UP} and δ_{DN} represent the incremental step caused by a single AE. For the scope of this demonstration,



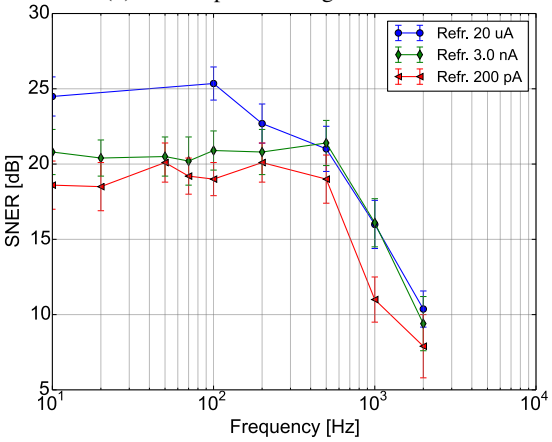
(a) ADC Delta Modulator



(b) Action potential signal reconstruction



(c) ADC Delta Modulator linearity



(d) ADC Delta Modulator refractory period

Fig. 6: a) ADC delta modulator thresholds, calibration, and signal reconstruction. b) action-potential signal reconstruction with refractory period set to 200 pA. c) Linearity of the converter over different input amplitudes. The frequency of the sine wave is set to 100 Hz. d) Signal to noise error ratio for different input sine wave frequencies. The amplitude of the input signal is set to 50 mV.

Result: *reconstructed*
while *incoming events* **do**
  $reconstructed(t) = reconstructed(t-1)$;
  if *channel events* $==$ *UP* **then**
    $reconstructed(t) = reconstructed(t-1) + \delta_{UP}$;
  end
  if *channel events* $==$ *DN* **then**
    $reconstructed(t) = reconstructed(t-1) - \delta_{DN}$;
  end
end

Algorithm 1: Signal reconstruction algorithm

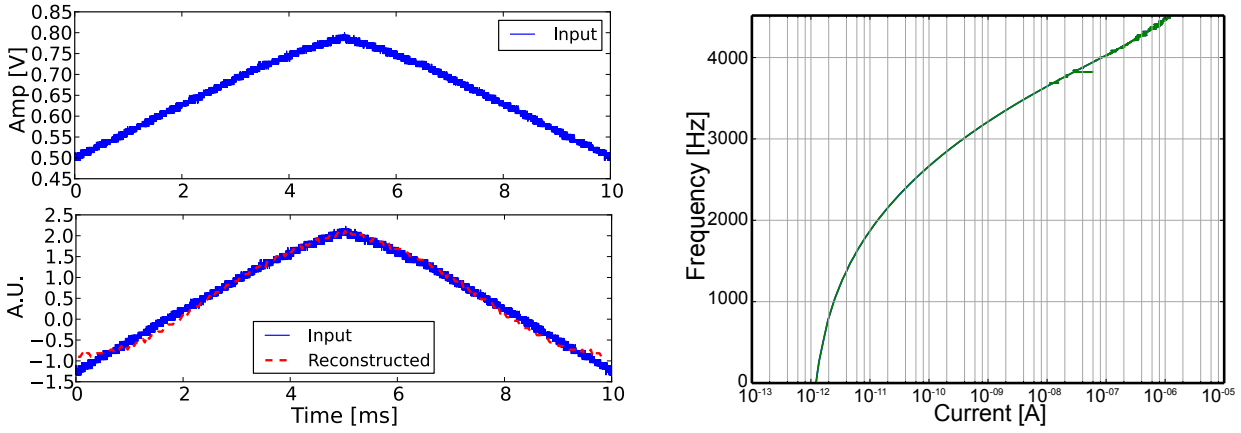
we calibrated the ADC delta modulator to output the same amount of spikes with a sinusoidal input, as shown in Fig. 6a.

Figure 6b shows a reconstruction of an action potential measurement. The top plot shows the action-potential signal after amplification with the LNA, superimposed to the reconstruction of the signal from the UP and DN events; the middle column shows the node V_{gout} of the asynchronous delta modulator that is constantly being compared with the thresholds; the bottom trace shows the UP and DN events. The signal to noise ratio between the two signals is $SNR = 9.4dB$.

In Fig. 6c we present a measure of the linearity of the ADC as a function of the amplitude of the input signal and of the delta step. The input signal used is a sine wave of 100 Hz. Changes in the δ step are reversely proportional to the total number of pulses, for a given input signal amplitude. In Fig. 6d we show the effect of the refractory period on the signal to noise error ratio. When the system is in the refractory period the input signal is discarded and for this fact the error accumulates.

C. Pulse ADC characterization

Figure 7 shows the input and output of the ADC pulse circuit for a linear sweep in the input voltage, in the range $0.5 < V_{in} < 0.8$ V. The line plot in the figure shows the reconstruction of the input signal. Figure 7b shows a measurement output frequency of the pulse ADC as a function of the input current. This curve is linear in the mid range of input currents, i.e., $10^{-10} < I_{in} < 10^{-6}$ A; while it has a clear non-linear drop

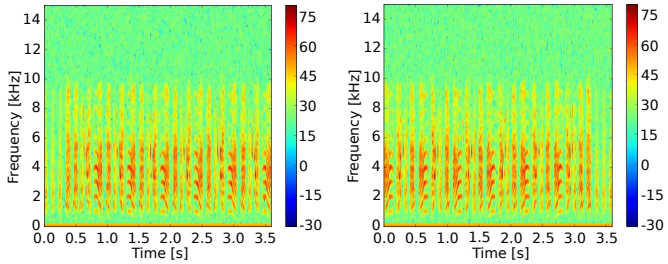


(a) (b)

Fig. 7: a) ADC Pulse recontruction. b) Output frequency versus input current.

Technology (CMOS)	0.18 μm 1P6M
Supply Voltage	1.8 V
Total area	0.178 mm^2
LNA power	90 μW
LNA RMS input-referred-noise	2.1 μV
A/D Delta Modulator power	100 Hz 55 μW
Bandwidth	$3Hz - 2.5kHz$
Gain (programmable)	54/50/45 dB
Effective NEF @ 45 dB	11

TABLE II: Neural Recording System characteristics



(a) (b)

Fig. 8: a) Auditory stimulus is bird's own song (BOS), class A. b) Auditory stimulus is reversed bird's own song (REV), class B.

in frequency for input currents below $I_{in} < 10^{-10}$ A. This drop is caused by the non-linearities of the input Differential Pair Integrator (DPI), used to convert V_{in} to the current signal I_d of transistor $MP2$ (see Fig. 1 Pulse ADC).

Signal reconstruction is carried by averaging the mean rate activity of the asynchronous digital output events. To obtain an absolute estimate of the input current, one has to interpolate the mean rate frequency with the characteristic visible in Fig. 7b.

D. Binary classification of neuro-biological recordings

1) Neural recording data and classification task: In this experiment we demonstrate the capabilities of the full system composed of the neural recording circuits and the neuromorphic processor. In this demonstration we use real neuro-biological recordings from Zebra Finches, a passerine bird from Central Australia. The data were taken in an anesthetized bird, in the lab of Prof. Hahnloser, and were kindly provided by Jenie Ondracek. The data consist of signals measured by four electrodes of auditory-forebrain neurons while natural sounds were being used as auditory stimuli. We focused on classifying two classes of auditory stimuli that have similar average energies, but different temporal structures: the bird's own song (BOS) (see Fig. 8a), and the reversed version of the bird's own song (REV) (see Fig. 8b). To this end, we grouped multiple channel recordings for the same class of auditory stimuli (BOS and REV respectively). The full data set consists of recordings from 229 neurons in the auditory forebrain of 16 birds. We used a subset of the data set, and we grouped recordings for six distinct birds. The auditory stimuli were presented in blocks; each stimulus block included consecutive repetitions (ranging from 18 to 60 times) of a fixed set of stimuli (BOS and REV). We used 17 stimuli blocks, and we had in total 2 groups of 32 recordings each with fifty stimulus repetitions. This was done in order to simulate a multi-electrode recording device with 32 parallel channels, whose output is projected to 64 silicon neurons as shown in Fig. 3 (32x2 as every channel is converted in UP and DN signals). Every group of recording was presented 28 times during the training phase and was presented 21 times during the testing phase. In Fig. 9a we show such grouping for 32 recordings in response to the BOS stimulus. The onset of the stimulus is aligned with the onset of the recordings. Figure 9b shows the output of the delta modulator for 80 ms. The reconstructed trace is shown in Fig. 9c. Even though the reconstruction does not faithfully reproduce the original recordings, the timing of all action potentials in the traces is well preserved.

2) Traning the neuromorphic classifier: Training the binary classifier is done by presenting multiple trials comprising signals obtained from grouped neural recordings, jointly with "teacher" signals generated externally. These teacher signals are Poisson spike trains that have a mean frequency of 25 Hz

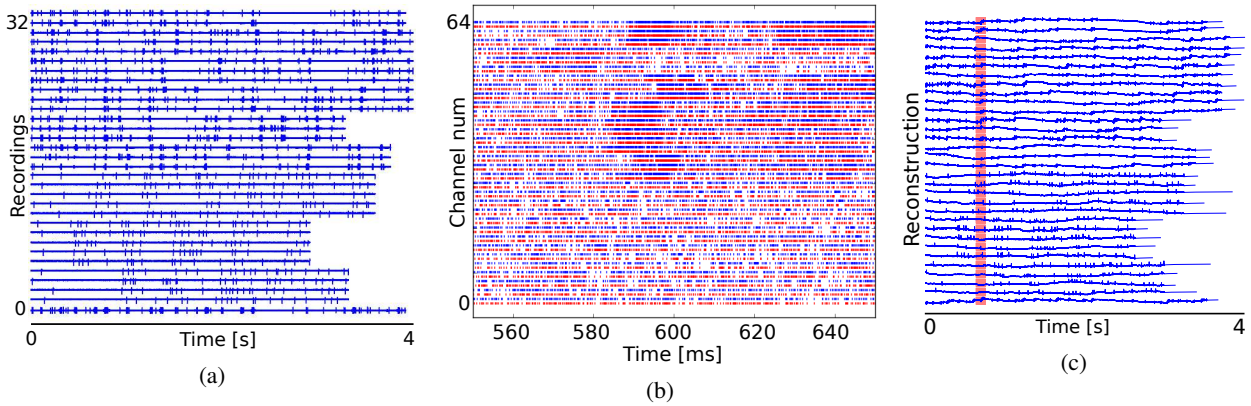


Fig. 9: a) Neural recordings aligned with stimulus onset. Auditory stimulus is (BOS). b) Recorded delta modulator output. Red is DN channel and blue is UP channel. c) Reconstructed recordings from Address-Event. Highlighted area shows the recording time of Fig. 9b, i.e.,≈ 80ms.

(teacher-false signal) or 150 Hz (teacher-true signal). These synthetic Poisson spike trains are directed to 4 AER virtual synapses which drive the neurons in the perceptron layer. The teacher-true signal drives the perceptron neurons such that they fire at a high rate (of approximately 100 Hz) when the true-class input stimulus is present, while the teacher-false signal drives neurons to fire at low frequency (about 5 Hz) when the false-class stimulus is applied. In this way, the synapses that connect the reservoir pool neurons to the perceptron layer neurons, driven by a teacher-true signal will tend to potentiate, making transitions to the high binary state; while synapses that connect reservoir neurons to the perceptron layer neurons driven by the teaching-false signal will tend to make transitions to the low binary state. At the beginning of the training session, all learning synapses are set to the low binary state. The training procedure of the classifier consists of presenting three repetitions of the same grouped recording trial, together with the appropriate teacher signal. In this condition a random subset of the plastic bi-stable synapses switch state, as prescribed by the stochastic spike-based learning rule described in Section II-G. This procedure is repeated over the entire set of training stimuli, for four different recording trials. At the end of the training procedure the state of the learning synapses is frozen, such that the pre and/or post-synaptic spikes will not cause any more transitions in the synaptic weights.

3) Testing the neuromorphic classifier: The testing phase is performed by presenting recordings from the same channels but for two different recording trials. The discrimination threshold, used to separate positive versus negative classification outputs is determined by re-playing all the teaching set to the classifier and by maximizing classifier-s performance. In the current example, the discrimination threshold was found to be at 5.1 Hz. During test trials, the spiking activity of the two pools of neurons is averaged and if it falls above/below the threshold value it is classified as a positive/negative sample. Figure 10 show examples of successful testing of the two distinct input patterns BOS and REV.

The overall performance of all testing trials is shown in

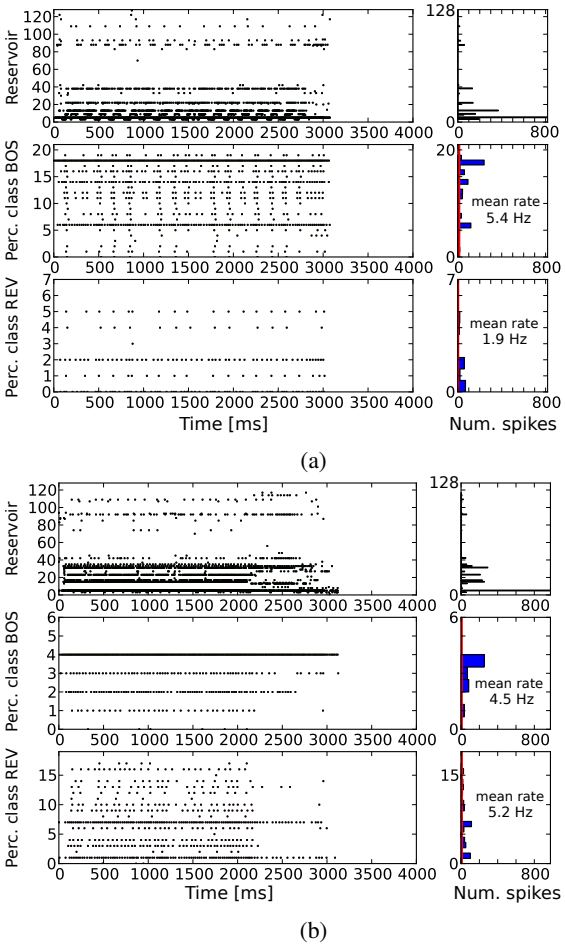


Fig. 10: 10a) Typical on-chip network activity during test trial: stimulus is BOS. Every dot in the plot represent a spike; The top panel shows reservoir activity; the middle panel shows activity of class A perceptrons; and the bottom panel shows class B perceptron activity. The plots in the right column represent histograms of spike counts during the 4 seconds of stimulation. 10b) Activity of a test trial in which the stimulus is REV.

Fig. 11. The final accuracy of this experiment is of approximately 96% on the test set: all the BOS stimuli were classified correctly and three REV stimuli were misclassified as false positives.

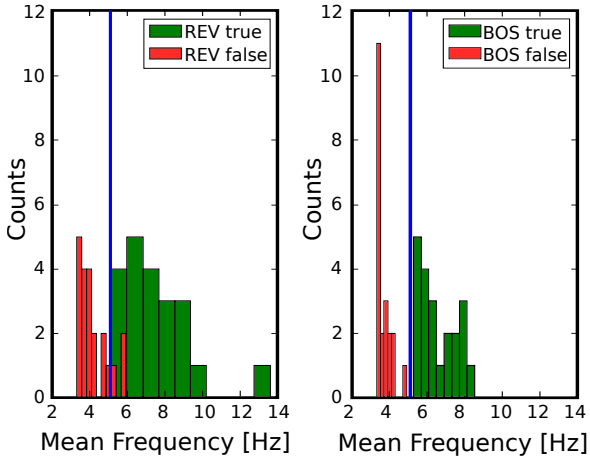


Fig. 11: Mean rate histograms of perceptrons activities during testing. The vertical line is the discrimination threshold. Three REV trials are wrongly classified as false positive.

IV. Discussions

Spiking neural networks have been shown to be able to carry out complex spatio-temporal processing and classification tasks [34], [35], [36], [37]. The specific topology of the neural network and the learning mechanisms adopted determine the signal processing ability of the system. These networks represent a promising solution for intelligent BMIs, as they can endow them with the ability to learn to decode neuronal signals into appropriate motor commands. In particular the reservoir computing paradigm is appealing for these types of BMIs because the recurrent population of spiking neurons acts as a pre-processing stage for the readout units, which increases the range of possible functions of the input streams that can be learned by linear combinations of the read-out units. The non-linearities present in the reservoir and the inhomogeneous properties of its computing elements are beneficial for this feature [34]. In BMIs the need of extracting a diverse pool of functions is directly related to a specific computational goal, such as predicting the movement of objects or non-linearly controlling a motor actuator. From the circuit implementation standpoint, the design choice of using asynchronous circuits is advantageous because the events generated from the recorded neural signals tend to be sparse. Therefore the encoding of the addresses in AE is a good solution in term of power consumption.

The mismatch and the limited precision of the analog circuits results in a diversity of responses that might be beneficial in a population coding scheme within the reservoir computing framework. This is also the case for the multi-perceptron network implemented in Section III-D, in which only few perceptrons responded to the trained stimulus. As in this population coding scheme every neuron was tuned to slightly different features, we adopted the "bagging" strategy to muse many weak classifiers in parallel to improve the overall classification performance [38], [39].

It has indeed been argued that neuromorphic electronic circuits offer a compact and ultra low-power substrate for implementing optimal on-line learning systems [15], as well as compact and power- and memory-efficient processing systems that are an attractive alternative to classical von Neumann architectures [40]. Input signals to these types of systems are typically provided by sending AER sequences of spikes from vision/auditory sensors, or produced on conventional computers. The circuits proposed in this paper can be used to create a new source of AEs that translate the activity of real neurons into the the relevant representation for further processing by neuromorphic computing cores. As recent developments are showing how to design neural networks that can optimally learn both feature extraction and pattern classification stages [41], in principle it is not necessary to detect individual action potentials in the neural recording data, or to carry out elaborate spike-sorting. The circuits presented here can lead to a new generation of compact, low-power, and adaptive BMIs that can be chronically implanted to that can carry out context-dependent learning for optimally driving and controlling prosthetic devices in real-world conditions.

V. Conclusions

We designed, fabricated, and tested a neural recording and processing system with AER interfacing circuits suitable for transmitting bio-potentials and LFP signals to neuromorphic computing architectures. Salient features of the system are reported in Table II.

We interfaced this system with a reconfigurable spiking neural network architecture, endowed with learning abilities. We exploited the parallel, distributed and low-power properties of the neuromorphic architecture to design a hardware reservoir computing framework, implemented as a recurrent neural network with fixed synaptic weights. Using this network we were able to optimally decode dynamic input signals and to configure it as a feature extraction layer capable of providing input to a second layer of on-chip perceptrons. The hardware perceptrons were trained to detect a specific sequence of activations of the recurrent units in a way to respond to a specific sequence of action potentials in the input signal, while ignoring other sequences. This work offers interesting prospectives for future intelligent neural-inspired BMI.

Acknowledgments

The results presented in this work were obtained also thanks to the collaboration of many colleagues, whom we like here to acknowledge: Qiao Ning, Juan Pablo Carbajal, David Bontrager, and Fabio Stefanini. The cricket auditory data used to demonstrate the potential capabilities of the systems described was kindly provided by Manfred Hartbauer. The data Zebra finch bird-song data, obtained in the lab of Prof. Hahnloser, was kindly provided by Jenie Ondracek. This work was supported by the SiCode EU project, under FET-Open grant

number FP7-284553, and by the neuroP EU project, under ERC grant number 257219. Some of the ideas and circuits presented in this work were inspired by the 2011 spring semester class project of the Neuromorphic Engineering 2 course, taught at the Institute of Neuroinformatics, University of Zurich and ETH Zurich.

REFERENCES

[1] T. Borghi, *et al.*, "An integrated low-noise multichannel system for neural signals amplification," *33rd European Solid State Circuits Conference, 2007. ESSCIRC*, pp. 456–459, Sept. 2007.

[2] R. Harrison, "The design of integrated circuits to observe brain activity," *Proceedings of the IEEE*, vol. 96, no. 7, pp. 1203–1216, 2008.

[3] S. Mandal and R. Sarpeshkar, "A bidirectional wireless link for neural prostheses that minimizes implanted power consumption," in *Biomedical Circuits and Systems Conference, (BioCAS), 2007*. IEEE, Nov. 2007, pp. 45–48.

[4] C. Charles, "Wireless data links for biomedical implants: Current research and future directions," in *Biomedical Circuits and Systems Conference, (BioCAS), 2007*. IEEE, Nov. 2007, pp. 13–16.

[5] A. Nurmikko, *et al.*, "Listening to brain microcircuits for interfacing with externalworld, progress in wireless implantable microelectronic neuroengineering devices," *Proceedings of the IEEE*, vol. 98, no. 3, pp. 375–388, 2010.

[6] B. K. D. Wise, *et al.*, "Microelectrodes, microelectronics, and implantable neural microsystems," *Proceedings of the IEEE*, vol. 96, no. 7, pp. 1184–1202, 2008.

[7] J. Aziz, *et al.*, "256-channel neural recording and delta compression microsystem with 3d electrodes," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 3, pp. 995–1005, March 2009.

[8] W. Wattanapanitch and R. Sarpeshkar, "A low-power 32-channel digitally programmable neural recording integrated circuit," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, no. 6, pp. 592–602, Dec 2011.

[9] I. Stevenson and K. Kording, "Dynamic causal models of neural system dynamics: current state and future extensions," *Nature Neuroscience*, vol. 14, no. 2, pp. 139–142, 2011.

[10] C. Lopez, *et al.*, "An implantable 455-active-electrode 52-channel CMOS neural probe," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 248–261, Jan 2014.

[11] J. Zhang, *et al.*, "A closed-loop compressive-sensing-based neural recording system," *Journal of neural engineering*, vol. 12, no. 3, p. 036005, 2015.

[12] S. Deiss, R. Douglas, and A. Whatley, "A pulse-coded communications infrastructure for neuromorphic systems," in *Pulsed Neural Networks*, W. Maass and C. Bishop, Eds. MIT Press, 1998, ch. 6, pp. 157–78.

[13] S.-C. Liu and T. Delbruck, "Neuromorphic sensory systems," *Current Opinion in Neurobiology*, vol. 20, no. 3, pp. 288–295, 2010.

[14] G. Indiveri, *et al.*, "Neuromorphic silicon neuron circuits," *Frontiers in Neuroscience*, vol. 5, pp. 1–23, 2011.

[15] E. Chicca, *et al.*, "Neuromorphic electronic circuits for building autonomous cognitive systems," *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1367–1388, Sep 2014.

[16] E. Neftci, *et al.*, "Synthesizing cognition in neuromorphic electronic systems," *Proceedings of the National Academy of Sciences*, vol. 110, no. 37, pp. E3468–E3476, 2013.

[17] R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 6, pp. 958–965, June 2003.

[18] W. Tang, *et al.*, "Continuous time level crossing sampling adc for biopotential recording systems," *IEEE transactions on circuits and systems. I, Regular papers: a publication of the IEEE Circuits and Systems Society*, vol. 60, no. 6, p. 1407, 2013.

[19] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128x128 120 dB 15 μs latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Feb 2008.

[20] T. Horiuchi, *et al.*, "A low-power CMOS neural amplifier with amplitude measurements for spike sorting," in *International Symposium on Circuits and Systems, (ISCAS), 2004*, vol. 4. IEEE, May 2004, pp. 29–32.

[21] ——, "Spike discrimination using amplitude measurements with a low-power cmos neural amplifier," in *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*. IEEE, 2007, pp. 3123–3126.

[22] S.-C. Liu, *et al.*, *Analog VLSI:Circuits and Principles*. MIT Press, 2002.

[23] N. Qiao, *et al.*, "A re-configurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses," *Frontiers in Neuroscience*, vol. 9, no. 141, 2015.

[24] H. Mostafa, *et al.*, "Automated synthesis of asynchronous event-based interfaces for neuromorphic systems," in *Circuit Theory and Design, (ECCTD) 2013 European Conference on*. IEEE, 2013, pp. 1–4.

[25] D. Barsakcioglu, *et al.*, "An analogue front-end model for developing neural spike sorting systems," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, pp. 216–227, 2014.

[26] J. Brader, W. Senn, and S. Fusi, "Learning real world stimuli in a neural network with spike-driven synaptic dynamics," *Neural Computation*, vol. 19, pp. 2881–2912, 2007.

[27] S. Mitra, S. Fusi, and G. Indiveri, "Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 3, no. 1, pp. 32–42, Feb. 2009.

[28] H. Mostafa, *et al.*, "A hybrid analog/digital spike-timing dependent plasticity learning circuit for neuromorphic VLSI multi-neuron architectures," in *International Symposium on Circuits and Systems, (ISCAS), 2014*. IEEE, 2014, pp. 854–857.

[29] F. Stefanini, *et al.*, "Pyncs: a microkernel for high-level definition and configuration of neuromorphic electronic systems," *Frontiers in Neuroinformatics*, vol. 8, no. 73, 2014.

[30] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, pp. 78–80, 2004.

[31] D. Han, *et al.*, "A 0.45 v 100-channel neural-recording ic with sub-/channel consumption in 0.18 cmos," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 7, no. 6, pp. 735–746, 2013.

[32] P. Kmon and P. Grybos, "Energy efficient low-noise multichannel neural amplifier in submicron cmos process," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 60, no. 7, pp. 1764–1775, July 2013.

[33] Y. Chen, *et al.*, "A digitally assisted, signal folding neural recording amplifier," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 8, no. 4, pp. 528–542, 2014.

[34] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, 2002.

[35] S. Sheik, *et al.*, "Spatio-temporal spike pattern classification in neuromorphic systems," in *Biomimetic and Biohybrid Systems*. Springer, 2013, pp. 262–273.

[36] M. Giulioni, *et al.*, "Classification of correlated patterns with a configurable analog VLSI neural network of spiking neurons and self-regulating plastic synapses," *Neural Computation*, vol. 21, no. 11, pp. 3106–3129, 2009.

[37] M. Giulioni, *et al.*, "Real time unsupervised learning of visual stimuli in neuromorphic VLSI systems," *Scientific Reports*, in press, 2015.

[38] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, 1996.

[39] M. Skurichina and R. P. Duin, "Bagging, boosting and the random subspace method for linear classifiers," *Pattern Analysis & Applications*, vol. 5, no. 2, pp. 121–135, 2002.

[40] G. Indiveri and S.-C. Liu, "Memory and information processing in neuromorphic systems," *Proceedings of the IEEE*, vol. 103, no. 8, pp. 1379–1397, 2015.

[41] G. Hinton, "Learning multiple layers of representation," *Trends in cognitive sciences*, vol. 11, no. 10, pp. 428–434, 2007.

Federico Corradi (S'13) received the B.Sc degree in Physics from Università degli studi di Parma, Italy, and the M.Sc. degree in Physics "cum laude" from La Sapienza University, Rome, Italy in 2010. In 2015 he received the Ph.D. degree in Natural Sciences from the University of Zürich (UZH) and the Ph.D. in Neuroscience from the Neuroscience Center Zürich (ZNZ). His research activities are at the interface between neuroscience and neuromorphic engineering. He is interested in understanding how biological neural circuits carry out computation and in mimicking the same style of computation in artificial systems. Currently he is employed as research and development scientist at INILabs, a spin-off company of the Institute of Neuroinformatics that produces and sells neuromorphic technologies.

Giacomo Indiveri (SM'03) is an Associate Professor at the Faculty of Science, University of Zürich, Switzerland. Indiveri received the M.Sc. degree in electrical engineering from the University of Genoa, Italy in 1992. Subsequently, he was awarded a doctoral postgraduate fellowship within the National Research and Training Program on "Technologies for Bioelectronics" from which he graduated with "summa cum laude" in 1995. He also obtained a Ph.D. degree in computer science and electrical engineering from the University of Genoa, Italy in 2004, and the "Habilitation" certificate in Neuromorphic Engineering from ETH Zurich, Switzerland in 2006. Indiveri carried out research on neuromorphic vision sensors as a Postdoctoral Research Fellow in the Division of Biology at the California Institute of Technology, Pasadena, CA, USA, and on neuromorphic selective attention systems as a postdoc at the Institute of Neuroinformatics of the University of Zurich and ETH Zurich, Switzerland. His current research interests lie in the study of real and artificial neural processing systems, and in the hardware implementation of neuromorphic cognitive systems, using full custom analog and digital VLSI technology. Indiveri is a member of several Technical Committees (TCs) of the IEEE Circuits and Systems society and a Fellow of the European Research Council